(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 May 2001 (17.05.2001)

PCT

(10) International Publication Number
WO 01/35393 A1

(51) International Patent Classification[7]: G10L 19/00

(21) International Application Number: PCT/GB00/04145

(22) International Filing Date: 26 October 2000 (26.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
99308858.2 8 November 1999 (08.11.1999) EP

(71) Applicant *(for all designated States except US)*: **BRITISH TELECOMMUNICATIONS public limited company** [GB/GB]; 81 Newgate Street, London EC1A 7AJ (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **GRAY, Philip** [GB/GB]; 15 Crabbe Street, Ipswich, Suffolk IP4 5HP (GB). **HOLLIER, Michael, Peter** [GB/GB]; "Casa Estepona", Westerfield Lane, Tuddenham, ST Martin IP6 9BH (GB).

(74) Agent: **LIDBETTER, Timothy, Guy, Edwin**; BT Group Legal Services, Intellectual Property Dept., 8th floor, Holborn Centre, 120 Holborn, London EC1N 2TE (GB).

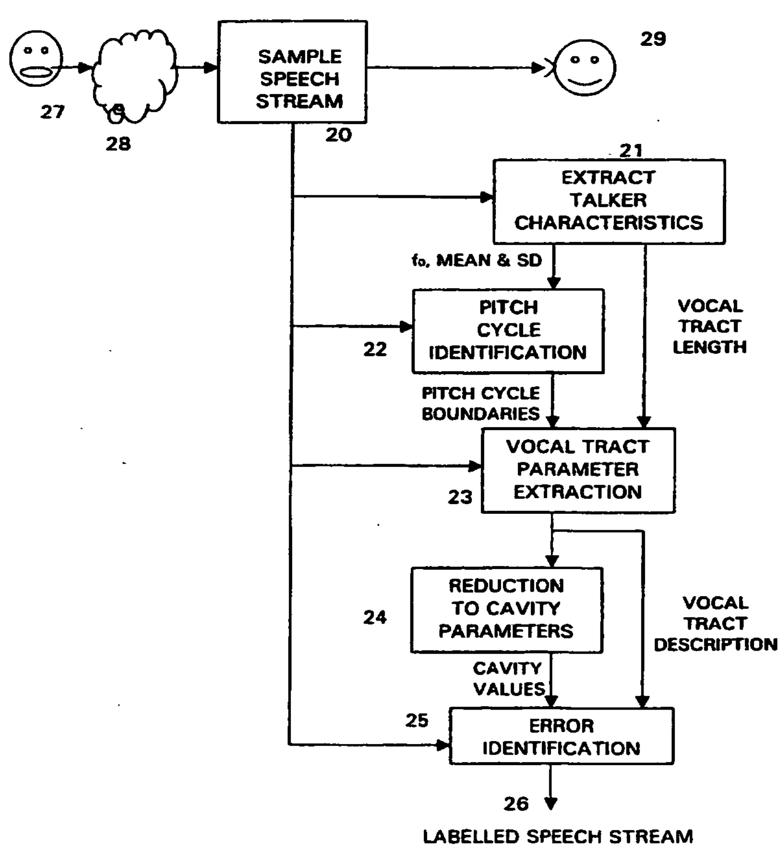
(81) Designated States *(national)*: AU, CA, JP, US.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published:
— *With international search report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NON-INTRUSIVE SPEECH-QUALITY ASSESSMENT



(57) Abstract: This invention relates to non-intrusive speech-quality assessment using vocal-tract models, in particular for testing telecommunications systems and equipment. This process requires the reduction of the speech stream under assessment into a set of parameters that are sensitive to the types of distortion to be assessed. Once parameterised, the data is used to generate a set of physiologically-based rules for error identification, using a parametric modelling of the shape of the vocal tract itself, by comparison between derived parameters and the output of models of physiologically realistic forms for the vocal tract, and the application of physical constraints on how these can change over time.

WO 01/35393 A1

NON-INTRUSIVE SPEECH-QUALITY ASSESSMENT

This invention relates to non-intrusive speech-quality assessment using vocal-tract models, in particular for testing telecommunications systems and equipment.

Customers are now able to choose a telecommunications service provider based upon price and quality of service. The decision is no longer fixed by monopolies or restricted by limited technology. A range of services is available with differing costs and quality of service. Service providers need the capability to predict customers' perceptions of quality so that networks can be optimised and maintained. Traditionally, networks have been characterised using linear assessment techniques, tone-based signals; and simple engineering metrics, such as signal-to-noise ratio. As networks become more complex, including non-linear elements such as echo cancellers and compressive speech coders, there is a requirement for assessment systems which bear a closer relationship to the human perception of signal quality. This role has typically been filled with expensive and time-consuming subjective tests using human subjects. Such tests are employed for commissioning new network elements, during the design of new coding algorithms, and for testing different network topologies.

Recent advances in perceptual modelling have led to the construction of objective auditory models, which can generate predictions of perceived telephony speech quality from a listener's perspective. These assessment techniques require a known test stimulus to excite a network connection and then use a perceptually-motivated comparison between a reference version of the known test stimulus, and a version of the same stimulus as degraded by the system under test, to provide a measure of the quality of the degraded version as it would be perceived by a human listener.

Figure 1 shows the principle of the BT Laboratories Perceptual Analysis Measurement System (PAMS), disclosed in International Patent Applications WO94/00922, WO95/01011, and WO95/15035. In this system the reference signal 11 comprises a speech-like test stimulus which is used to excite the connection under test 10 to generate a degraded signal 12. The two signals are then compared in the analysis process 1 to generate an output 18 indicative of the subjective impact of the degradation of the signal 12 when compared with the reference signal 11.

ISDOCID: <WO___0135393A1_I_>

Such assessment techniques are known as "intrusive" because they require the withdrawal of the connection under test 10 from normal service so that it can be excited with a known test stimulus 11. Removing a connection from normal service renders it unavailable to customers and is expensive to the service provider. In addition, the conditions that generate distortions and errors could be due to network loading levels that are only present at peak times. An out-of-hours assessment could therefore generate artificial quality scores. This means that reliable intrusive testing is relatively expensive in terms of capacity on a customer's network connection.

In general, it would be preferable to continuously monitor the quality of speech at a particular point in the network. In this case, a "non-intrusive" solution is attractive, utilising the in-service signal to make predictions of quality. Given this information, network traffic can be re-routed through less congested parts of the network if quality drops.

A fundamentally different approach is required to analyse a degraded speech signal without a reference signal. The entire process takes place "downstream" of the equipment under test. Non-intrusive techniques are discussed in International Patent Specifications WO96/06495 and WO96/06496. Current non-intrusive assessment equipment performs measurements such as echo, delay, noise and loudness in an attempt to predict the clarity of a connection. However, a customer's perception of speech quality is also affected by distortions and irregularities in the speech structure, which are not described by such simple measures.

International Patent Specification WO97/05730 describes a system of this general type which aims to generate an output indicative of how plausible it is that the passing audio stream was generated by the human vocal production system. This is achieved by comparing the audio stream with a spectral model representative of the sounds capable of production by the human vocal system. This process requires pattern recognition to distinguish the spectral characteristics representative of speech and of distortion, so that their presence can be identified.

These analysis processes use spectral models, although physiological models have previously been used for speech synthesis – see for example the use of each types of model for these respective purposes in International patent specifications WO96/06496 and WO97/00432. Unlike a physiological model, spectral models are empirical, and have no intrinsic basis on which to identify what sounds the vocal

3DOCID: <WO___0135393A1_I_>

tract is capable of producing. However, the physiological articulatory models used in the synthesis of continuous speech utilise constraints to ensure the generated speech is smooth and natural sounding. These models would therefore be unsuitable for an assessment process, since in such a process the parameters generated must also be capable of representing "illegal" vocal-tract shapes that the constraints used by such a synthesis model would ordinarily remove. It is the regions that are in error or distorted that contain the information for such an assessment; to remove this at the parameterisation stage would make a subsequent analysis of their properties redundant.

According to the present invention, there is provided a method of identifying distortion in a signal carrying speech, in which the signal is analysed according to parameters derived from a set of physiologically-based rules using a parametric model of the human vocal tract, to identify parts of the signal which could not have been generated by the human vocal tract. This differs from the prior art systems described above which use empirical spectral analysis rules to distinguish speech fromother signals. The analysis process used in the invention instead considers whether physiological combinations exist that could generate a given sound, in order to determine whether that sound should be identified as possible to have been formed by a human vocal tract.

Preferably the analysis process comprises the step of reducing a speech stream into a set of parameters that are sensitive to the types of distortion to be assessed.

Cavity tracking techniques and context based error spotting may be used to identify signal errors. This allows both instantaneous abnormalities and sequential errors to be identified. Articulatory control parameters (parameters derived from the movement of the individual muscles which control the vocal tract) are extremely useful for speech synthesis applications where their direct relationships with the speech production system can be exploited. However, they are difficult to use for analysis, because the articulatory control parameters are heavily constrained to maintain their conformance to the production of real vocal tract configurations. It is therefore difficult to model error conditions, which necessarily require the modelling of conditions that the vocal tract cannot produce. It is therefore preferred to use acoustic tube models. Such models allow the derivation of vocal-tract descriptors

ISDOCID: <WO___0135393A1_I_>

directly from the speech waveform, which is attractive for the present analysis problem, as physiologically unlikely conditions are readily identifiable.

An embodiment of the invention will now be described, with reference to the accompanying drawings, in which

Figure 1 is a schematic illustration of the PAMS intrusive assessment system already discussed.

Figure 2 is a schematic illustration of the system according to the invention

Figure 3 illustrates the use of a variable frame length;

Figure 4 is an illustration of the pitch boundaries of a voiced speech event.

Figure 5 illustrates a simplified uniform-cross-sectional-area tube model used in the invention.

Figure 6 is an illustration of the human vocal tract.

Figure 7 illustrates a cavity area sequence;

Non-intrusive speech quality assessment processes require parameters with specific properties to be extracted from the speech stream. They should be sensitive to the types of distortions that occur in the network under test; they should be consistent across talkers; and they should not generate ambiguous mappings between speech events and parameters.

Figure 2 shows illustratively the steps carried out by the process of the invention. It will be understood that these may be carried out by software controlling a general-purpose computer. The signal generated by a talker 27 degraded by the system 28 under test. It is sampled at point 20 and concurrently transmitted to the end user 29. The parameters and characteristics identified from the process are used to generate an output 26 indicative of the subjective impact of the degradation of the signal 2, compared with the signal assumed to have been supplied by the source 27 to the system 28 under test.

The degraded signal 2 is first sampled (step 20), and several individual processes are then carried out on the sampled signal.

A major problem with non-intrusive speech-quality assessment is lack of information concerning talker characteristics. In the laboratory it is possible to generate talker-specific algorithms with near-perfect error spotting capabilities. These work well because prior knowledge of the talker has been used in development, even though no reference was used. In the real world operation with multiple talkers is

SDOCID: <WO___0135393A1_I_>

necessary, and individual talker variation can generate significant performance reductions.

The process of the present invention compensates for this type of error by including talker characteristics in both the parameterisation stage and also the assessment phase of the algorithm. The talker characteristics are restricted to those that can be derived from the speech waveform itself, but still yield performance improvements.

A model is used in which the overall shape of the human vocal tract is described for each pitch cycle. This approach assumes that the speech to be analysed is voiced, (i.e the vocal chords are vibrating, for example vowel sounds) so that the driving stimulus can be assumed to be impulsive. The vocal characteristics of the individual talker 27 are first identified (process 21). These are features that are invariant for that talker 27, such as the average fundamental frequency f_0 of the voice, which depends on the length of the vocal tract. This process 21 is carried out as follows. It uses a section of speech in the order of 10 seconds to characterise the talker by extracting information about the fundamental frequency and the third formant (third harmonic) values. These values are calculated for the voiced sections of speech only. The mean and standard deviation of the fundamental frequency is used later, during the pitch-cycle identification. The mean of the third formant values is used to estimate the length of the vocal tract.

The number of tubes used to calculate the cross sectional areas should be related to the length of the talkers vocal-tract, measured (as deviations from a notional figure of 17cm) according to information from the formant positions within the speech waveform. Using the third formant, which is generally present with telephony bandwidth restrictions, it is possible to alter the number of tubes to populate the equivalent lossless tube model.

The appropriate number of tube sections is given by the closest integer value to N_l, where: $$N_l = 2lf_s/c$$

where: l = vocal tract length; f_s = sample frequency; c = speed of sound: (330m/sec).

Assuming a sampling frequency of 16kHz, for the average talker of vocal tract length 17cm and average 3[rd] formant frequency of 2500Hz, this leads to sixteen cross-sectional areas being required to populate the tube model. Using a direct proportionality between the average 3[rd] formant frequency for a talker and the

ISDOCID: <WO___0135393A1_I_>

length of the vocal tract it is possible to estimate the value of l in the equation above: this estimated value l_m is calculated from:

$$l_m/17 = 2500/d$$

where d, average 3[rd] formant value

For a female talker with an average third formant frequency of 3kHz, this gives an estimated vocal tract length of 14cm, and the number of tube sections N_t as fourteen.

This method for vocal tract length normalisation reduces the variation in the parameters extracted from the speech stream so that a general set of error identification rules can be used which are not affected by variations between talker, of which pitch is the main concern.

Once characteristion has been carried out using the initial ten second section of speech, the parameters identified (mean fundamental frequency, standard deviation, and vocal tract length) may be used for the rest of the speech stream, periodically repeating the initial process in order to detect changes in the talker 27.

The samples taken from the signal 2 (step 20) are next used to generate speech parameters from these characteristics. An initial stage of pitch synchronisation is carried out (step 22). This stage generates a pitch-labelled speech stream, enabling the extraction of parameters from the voiced sections of speech on a variable time base. This allows synchronisation with the speech waveform production system, namely the human speech organs, allowing parameters to be derived from whole pitch-periods. This is achieved by selecting the number of samples in each frame such that the frame length corresponds with a cycle of the talker's speech, as shown in Figure 3. Thus, if the talker's speech rises and falls in pitch the frame length will track it. This reduces the dependence of the parameterisation on gross physical talker properties such as their average fundamental frequency. Note that the actual sampling rate carried out in the sampling step 20 remains constant at 16kHz – it is the number of such samples going to make up each frame which is varied.

SDOCID: <WO___0135393A1_I_>

Various methods exist for the generation of pitch-synchronous boundaries for parameterisation. The present embodiment uses a hybrid temporal spectral method, as described by the inventors in their paper *"Constraint-based pitch-cycle identification using a hybrid temporal spectral method" – 105th AES Convention, 1998*. This process uses the mean fundamental frequency f_0, and the standard deviation of this value, to constrain the search for these boundaries

The output of this non-real time method can be seen in Figure 4 , which shows the pitch boundaries (marked "X") for a voiced speech event. It can be seen that these are synchronised with the largest peaks in the voice signal, and thus occur at the same frequency as the fundamental frequency of the talker's voice. The lengths of the pitch cycles vary to track changes in the pitch of the talker's voice.

Having identified the pitch-synchronous parameters, the parameterisation of the vocal tract can now be done (step 23). It is important that no constraints are imposed during the parameterisation stages that could smooth out or remove signal errors, as they would then not be available for identification in the error identification stage. Articulatory models used in the synthesis of continuous speech utilise constraints to ensure the generated speech is smooth and natural sounding. The parameters generated by a non-intrusive assessment must be capable of representing illegal vocal-tract shapes that would ordinarily be removed by constraints if a synthesis model were used. It is the regions that are in error or distorted that contain the information for such an assessment, to remove this at the parameterisation stage would make a subsequent analysis of their properties redundant.

In the process of the present embodiment, reflection coefficients are first calculated directly from the speech waveform over the period of a pitch cycle, and these are used to determine the magnitude of each change in cross section area of the vocal tract model, using the number of individual tube elements derived from the talker characteristics already derived (step 21). The diameters of the tubes to be used in the model can then be derived from these boundary conditions (step 23). An illustration of this representation can be seen in Figure 5, which shows a simplified uniform-cross-sectional-area model of a vocal tract. In this model the vocal tract is modelled as a series of cylindrical tubes having uniform length, and having individual cross sectional areas selected to correspond with the various parts of the vocal tract. The number of such tubes was determined in the preliminary step 21.

ISDOCID: <WO___0135393A1_I_>

For comparison, the true shape of the human vocal tract is illustrated in Figure 6. In the left part of Figure 6 there is shown a cross section of a side view of the lower head and throat, with six section lines numbered 1 to 6. In the right part of Figure 6 are shown the views taken on these section lines. The non-circular shape of the real vocal tract, and the fact that the real transitions are not abrupt steps, result in higher harmonics being modelled less well in the tube model of Figure 5 , but these do not affect the analysis for present purposes. We can therefore use a uniform-cross-sectional-area tube model to describe the instantaneous state of the vocal tract.

Certain errors may be apparent from the individual vocal tract parameters themselves, and can be identified directly. However, more generalised error identification rules may be derived from parameters derived by aggregating these terms. For this reason, the dimensionality of the vocal-tract description is reduced even further at this point to maintain a constant number (step 24). Methods that track constrictions within the tract yield large variations in the individual cavity parameters during steady-state clean speech attributable to minor differences in the calculation of the constriction point. These differences are significant enough to mask certain errors in degraded speech streams.

It has been found experimentally that the best results are produced by splitting the tract into three regions: front cavity, rear cavity, and jaw opening. The accompanying table shows the number of tube elements making up each of the three cavities for each of the numbers of tubes considered.

Total Number of Tubes	Rear Cavity	Front Cavity	Jaw Opening
12	5	5	2
13	5	6	2
14	6	5	3
15	6	6	3
16	7	6	3
17	7	7	3
18	8	7	3

The total cross sectional area in each of the tube subsets is aggregated to give an indication of cavity opening in each case.

Examples of cavity traces can be seen in Figure 7, showing (in the lower part of the figure) the variation in area in each of the three defined cavities during the

SDOCID: <WO___0135393A1_I_>

passage of speech "He was genuinely sorry to see them go", whose analogue representation is indicated in the part of the Figure. The blank sections correspond to unvoiced sounds and silences, which are not modelled using this system. This is because the cross sectional area parameters can only be calculated during a pitched voice event, such as those which involve glottal excitation caused by vibration of the vocal chords. Under these conditions parameters can be extracted from the speech waveform which describes its state. The rest of the events are unvoiced and are caused by constrictions at different places in the tract causing turbulent airflow, or even a complete closure. The state of the articulators is not so easy to estimate for such events.

The cavity sizes extracted (step 24) from the vocal tract parameters for each pitch frame are next assessed for physiological violations (step 25). Any such violations are taken to be caused by degradation of the signal 2, and cause an error to be identified. These errors are identified in the output 26. Errors can be categorised in two major classes, instantaneous and sequential.

Instantaneous errors are identified where the size of the cavity value at a given instance in time is assessed as implying a shape that would be impossible for a human vocal tract to take. An extreme example of this is that certain signal distortions can yield excessively large apparent jaw openings – for example 30 cm, and could not have been produced by a human vocal tract. There are other more subtle situations, which have been found empirically, where certain *combinations* of cavity sizes do not occur in human speech. Any such physiological impossibilities are labelled accordingly, as being indicative of a signal distortion.

One of the most common areas of degradation in speech streams in the modern telephony network is through speech coding. Specialised coding schemes, specific to voice signals, can generate distortions when incorrect outputs are generated from the coded parameter stream. In this situation the individual frames may seem entirely appropriate when viewed in isolation, but when the properties of the adjacent frames are taken into account, an error in the degraded signal is apparent. These types of distortion have been termed *"sequential errors"*. Sequential errors occur quite often in heavily coded speech streams. If incorrect parameters arrive at the decoder, because of miscoding or corruption during transmission, the reconstructed speech stream may contain a spurious speech event. This event may

INSDOCID: <WO___0135393A1_I_>

be "legal" – that is, if viewed in isolation or over a short time period it does not require a physiologically impossible instantaneous configuration of the vocal tract - but when heard would be an obvious that an error was present. These types of distortion are identified in the error identification step by assessing the sizes of cavities and vocal tract parameters, in conjunction with the values for preceding and subsequent frames, to identify sequences of cavity sizes which are indicative of signal distortion.

The error identification process 25 is operates according to predetermined rules arranged to identify individual cavity values, or sequences of such values, which cannot occur physiologically. Some speech events are capable of generation by more than configuration of the vocal tract. This may result in apparent sequential errors when the process responds to a sequence including such an event, if the process selects a vocal tract configuration different from that actually used by the talker. The process is arranged to identify any apparent sequential errors which could result from such ambiguities, so that it can avoid mislabelling them as errors.

SDOCID: <WO___0135393A1_I_>

CLAIMS

1. A method of identifying distortion in a signal carrying speech, in which the signal is analysed according to parameters derived from a set of physiologically-based rules using a parametric model of the human vocal tract, to identify parts of the signal which could not have been generated by the human vocal tract.

2. A method according to claim 1, in which the analysis of the signal comprises identification of the instantaneous configuration of the parametric model.

3. A method according to claim 1 or 2 in which the analysis of the signal comprises the analysis of sequences of configurations of the parametric model

4. A method according to any preceding claim, in which cavity tracking and context based error spotting are used to identify signal errors.

5. A method according to claim 4, in which the parametric model comprises a series of cylindrical tubes, the dimensions of the tubes being derived from reflection coefficients determined from analysis of the original signal.

6. A method according to claim 5, wherein the number of tubes in the series is determined from a preliminary analysis of the signal to identify vocal characteristics characteristic of the talker generating the signal.

7. A method according to any preceding claim, in which pitch-synchronised frames are selected for analysis.

8. Apparatus for assessing the quality of a signal carrying speech, comprising processing means for performing the method of any of the preceding claims.

9. A data carrier, carrying program data for programming a computer to perform the method of any of claims 1 to 8.

ISDOCID: <WO___0135393A1_I_>

10. Apparatus for assessing the quality of a signal carrying speech, comprising means for deriving parameters of the signal from a set of physiologically-based rules using a parametric model of the human vocal tract, and identifying parameters which indicate whether the signal could have been generated by the human vocal tract.

11. Apparatus according to claim 10, comprising means for identification of the instantaneous configuration of the parametric model.

12. Apparatus according to claim 10 or 11 comprising means for analysis of sequences of configurations of the parametric model

13. Apparatus method according to claim 10, 11 or 12, wherein the parameter-deriving means include cavity tracking means and context based error spotting means.

14. Apparatus according to claim 13, comprsising means for analysis of the original signal to identify relection coefficients, and model generation means for generation of a parametric model comprising a series of cylindrical tubes, the dimensions of the tubes being derived from the reflection coefficients.

15. Apparatus according to claim 14, comprising means for making a preliminary analysis of the signal to identify vocal characteristics characteristic of the talker generating the signal, and wherein the parameteric model generation means is arranged to select the number of tubes in the series according to the said vocal characteristics.

16. Apparatus method according to claim 10, 11, 12, 13, 14 or 15, in which tye analysis meamns is arranged to select pitch-synchronised frames.

3DOCID: <WO___0135393A1_I_>

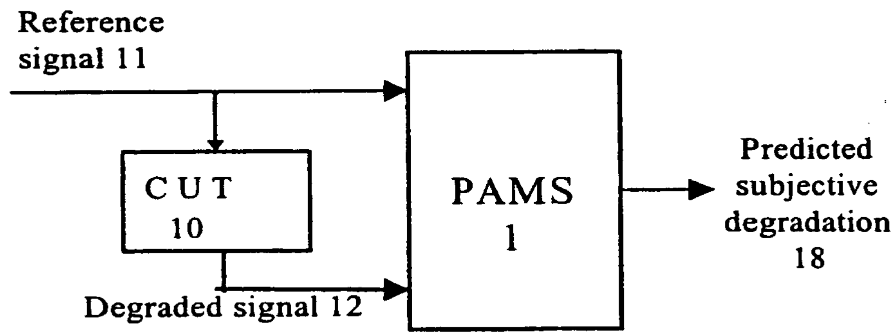
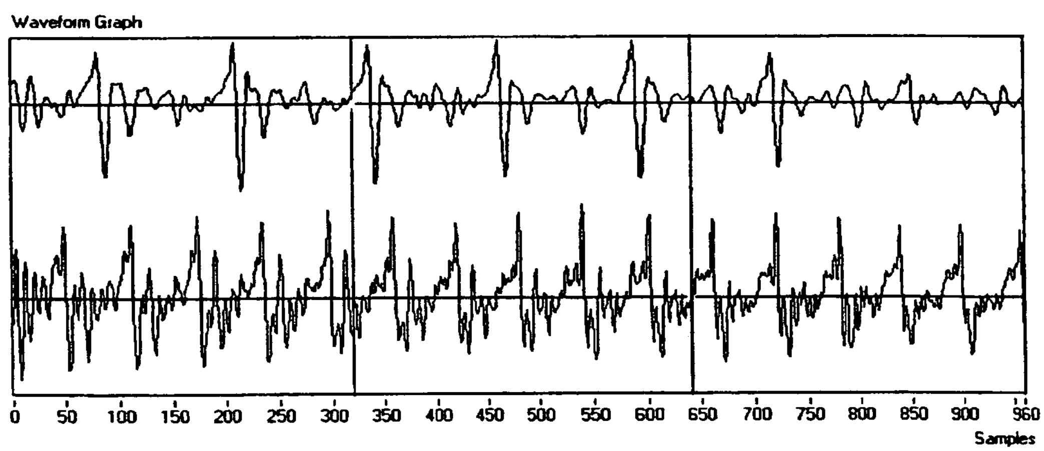


Figure 1

Figure 3



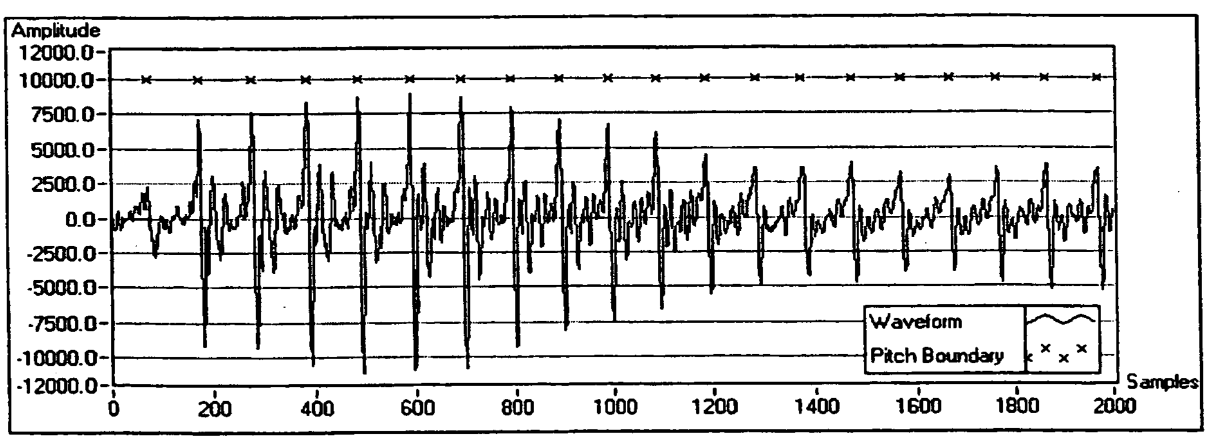
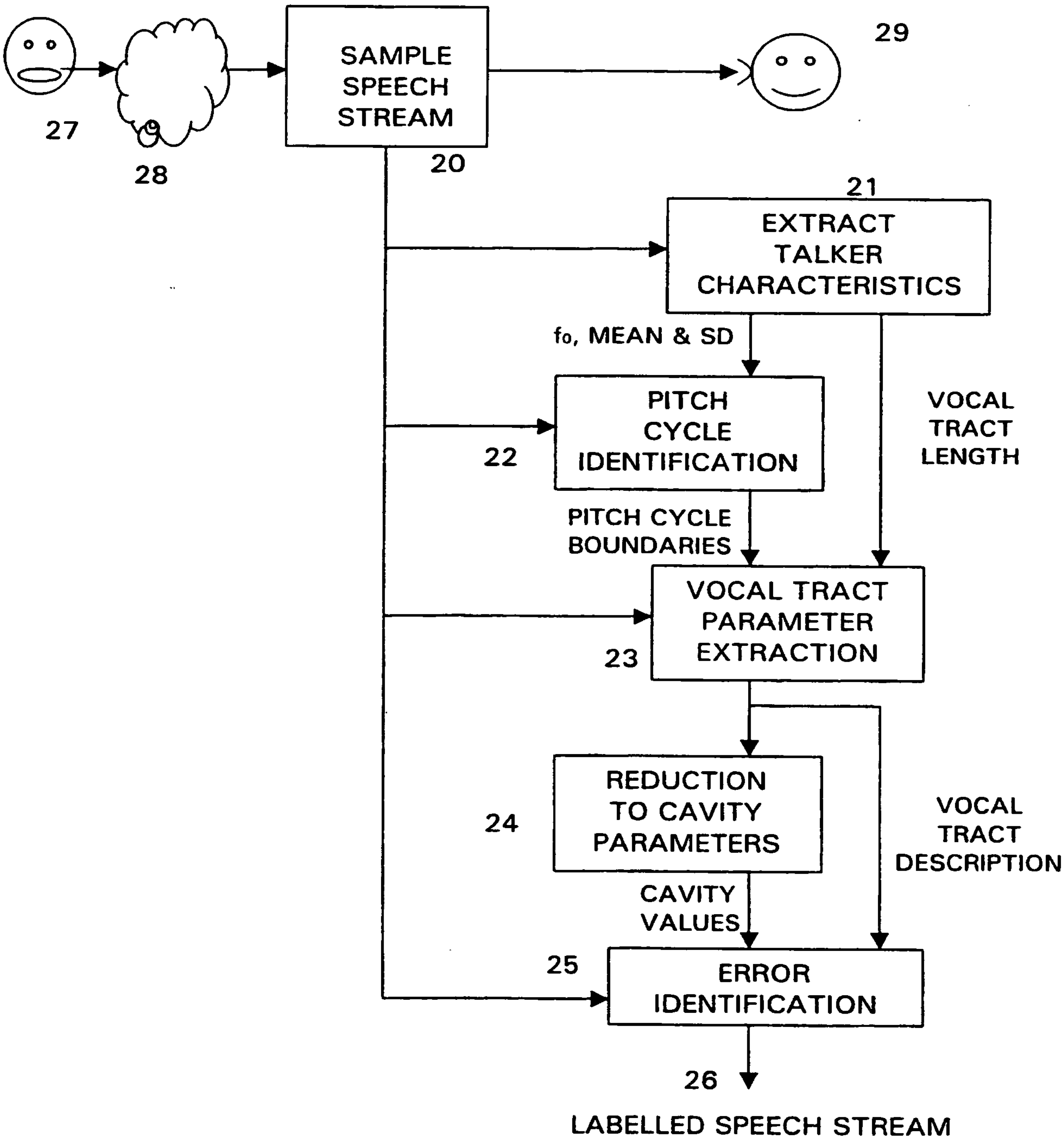


Figure 4

ISDOCID: <WO___0135393A1_I_>

Figure 2



DOCID: <WO___0135393A1_I_>

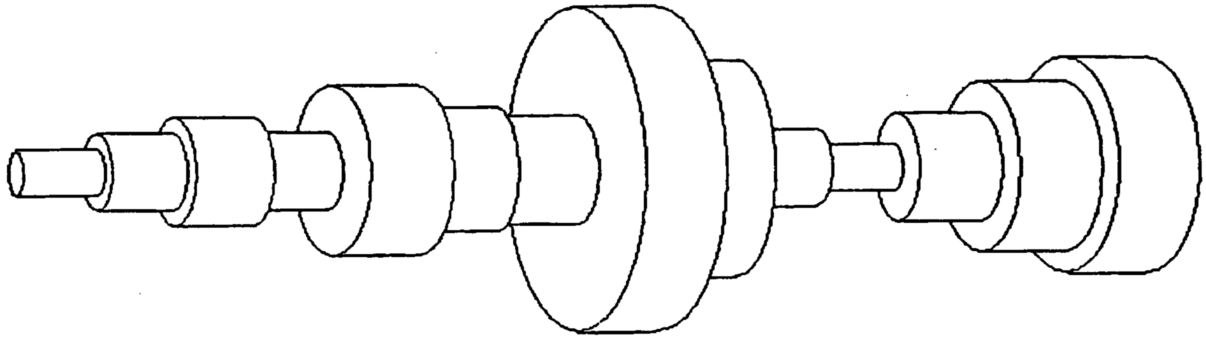
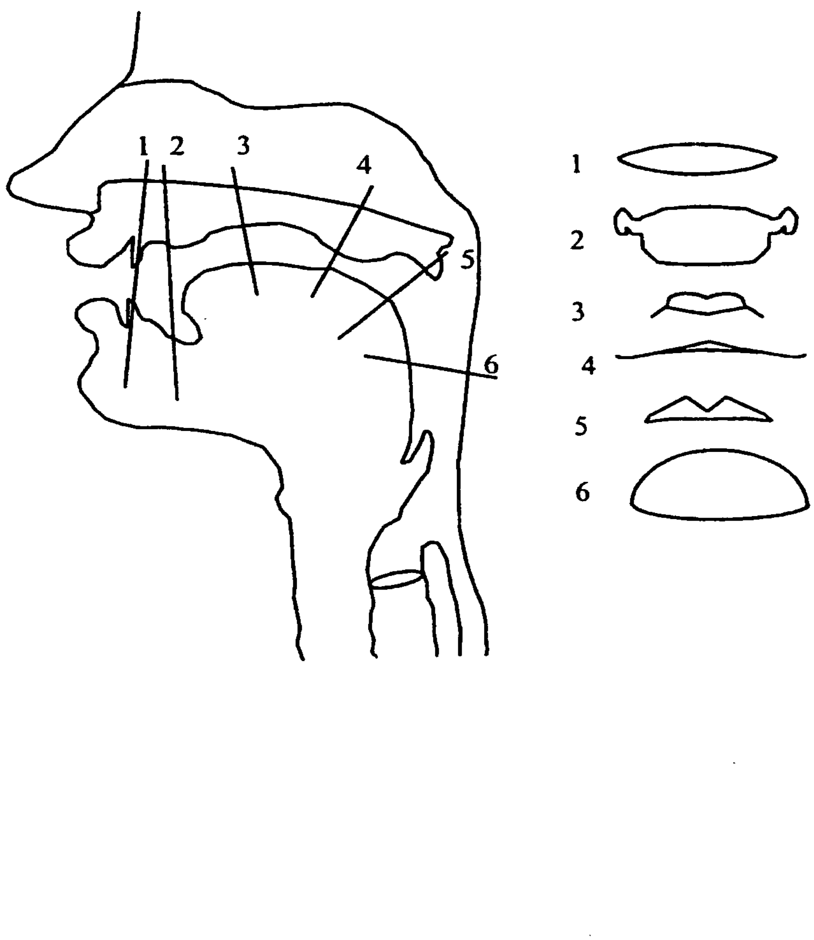


Figure 5

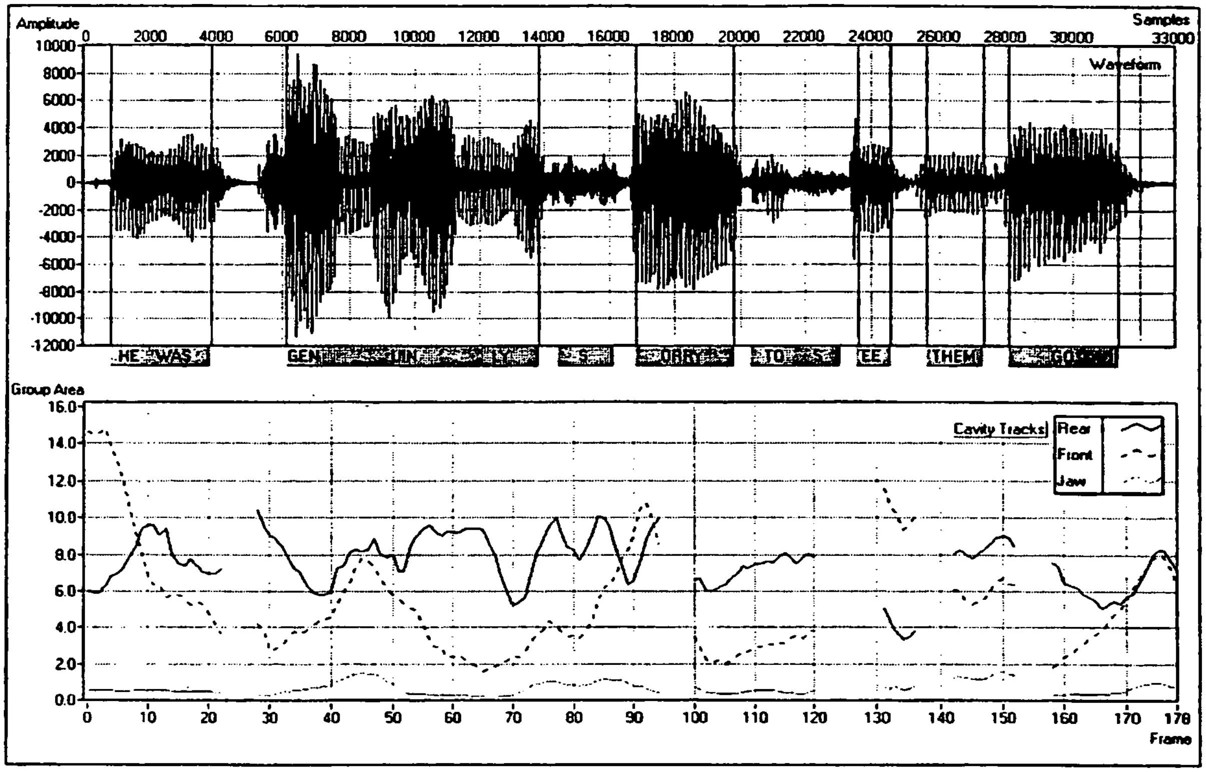
Figure 6



NSDOCID: <WO___0135393A1_I_>

BEST AVAILABLE COPY

Figure 7



SDOCID: <WO___0135393A1_I_>

INTERNATIONAL SEARCH REPORT

Intern. ,nal Application No
PCT/GB 00/04145

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G10L19/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 97 05730 A (HOLLIER MICHAEL PETER ;GRAY PHILIP (GB); BRITISH TELECOMM (GB); SH) 13 February 1997 (1997-02-13) cited in the application abstract; figure 2 page 1, line 3-5 page 2, line 3-8 page 3, line 10-18 page 3, line 25 -page 4, line 1 page 4, line 8-15 & US 6 035 270 A 7 March 2000 (2000-03-07) ---	1-3,8-12

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 January 2001	17/01/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Quélavoine, R

1

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	MRAYATI M ET AL: "THE ACOUSTIC-AREA FUNCTION INVERSION PROBLEM AND THE DISTINCTIVE REGION MODEL" PROCEEDINGS OF THE EUROPEAN SIGNAL PROCESSING CONFERENCE (EUSIPCO),NL,AMSTERDAM, ELSEVIER, vol. CONF. 6, 1992, pages 155-158, XP000348636 ISBN: 0-444-89587-6 * page 155, section "1. introduction" *	5,6,14, 15
A	COETZEE H J ET AL: "AN LSP BASED SPEECH QUALITY MEASURE" INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH & SIGNAL PROCESSING. ICASSP,US,NEW YORK, IEEE, vol. CONF. 14, 23 May 1989 (1989-05-23), pages 596-599, XP000089793 abstract	1-16

1

3DOCID: <WO___0135393A1_I_>

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/GB 00/04145

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9705730 A	13-02-1997	AU 694932 B	06-08-1998
		AU 6623296 A	26-02-1997
		CA 2225407 A	13-02-1997
		CN 1192309 A	02-09-1998
		EP 0840975 A	13-05-1998
		JP 11510334 T	07-09-1999
		NO 980331 A	26-01-1998
		NZ 313705 A	25-11-1998
		US 6035270 A	07-03-2000
		AU 711615 B	14-10-1999
		AU 1553197 A	16-09-1997
		CA 2237814 A	04-09-1997
		EP 0879527 A	25-11-1998
		WO 9732428 A	04-09-1997
		JP 2000506327 T	23-05-2000
		NO 983959 A	28-08-1998
		US 6119083 A	12-09-2000
		US 5799133 A	25-08-1998